



①9 BUNDESREPUBLIK  
DEUTSCHLAND



DEUTSCHES  
PATENTAMT

⑫ **Offenlegungsschrift**  
⑩ **DE 43 25 404 A 1**

⑤1 Int. Cl.<sup>6</sup>:  
**G 10 L 5/06**

②1 Aktenzeichen: P 43 25 404.7  
②2 Anmeldetag: 29. 7. 93  
④3 Offenlegungstag: 2. 2. 95

DE 43 25 404 A 1

⑦1 Anmelder:  
Telenorma GmbH, 60326 Frankfurt, DE

⑦2 Erfinder:  
Falkhausen, Markus, Dipl.-Phys., 63071 Offenbach, DE;  
Nicol, Natascha, Dipl.-Phys., 60316 Frankfurt, DE;  
Reininger, Herbert, Dr., 61194 Niddatal, DE;  
Wolf, Dietrich, Prof. Dr., 63322 Rödermark, DE;  
Zinke, Joachim, Dr., 61194 Niddatal, DE

⑤4 Verfahren zum Ermitteln und Klassifizieren von Störgeräuschtypen

⑤7 Mit einer für die Spracherkennung geeigneten Anordnung soll bei relativ geringem Speicher- und Rechneraufwand zunächst ein Geräusch eindeutig oder annähernd bestimmt werden. Das dabei erzielte Ergebnis soll zur Optimierung einer Spracherkennung oder zur Erzeugung von Schaltbefehlen dienen.

Es wird zunächst ein Standardcodebuch erstellt, welches die Referenzmuster von Sprachäußerungen ohne Störgeräusch aufnimmt. Danach werden Sprachäußerungen zusammen mit verschiedenen Störgeräuschtypen einzeln trainiert, wobei jeweils Merkmalsvektoren erzeugt werden, die mit im Standardcodebuch vorhandenen Merkmalsvektoren verglichen werden. Die dabei entstehenden Häufigkeitsverteilungen der Merkmalsvektoren werden in einem Indexspeicher abgelegt. Die im Indexspeicher abgelegten Häufigkeitsverteilungen der Indizes von Merkmalsvektoren werden dann mit den Indizes von Merkmalsvektoren aus der aktuell eingegebenen, mit einem Störgeräuschtyp behafteten Sprachäußerung verglichen. Wenn sich dabei ein Extremwert ergibt, so ist ein Störgeräuschtyp gefunden worden. Das dabei entstehende Signal kann zu einer genaueren Spracherkennung oder zur Erzeugung von Schaltbefehlen dienen.

DE 43 25 404 A 1

Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen

BUNDESDRUCKEREI 11. 94 408 065/208

6/29

Die Erfindung betrifft ein Verfahren zum Ermitteln und Klassifizieren von Störgeräushtypen unter Anwendung von für die Sprachanalyse bekannten Verfahren nach dem Oberbegriff des Patentanspruchs 1.

In einem Fachbericht mit dem Titel "Einzelwort-Erkennung in gerauschvoller Umgebung", der in einem Tagungsband zur Konferenz über Sprachverarbeitung unter ungünstigen Bedingungen (Cannes-Mandelieu, vom 10. bis 13. Nov. 1992; ISSN 1018-4554) ab Seite 215 abgedruckt ist, wird ein Verfahren beschrieben, womit eine Spracherkennung auch dann ermöglicht wird, wenn Geräusche auftreten. Unter Anwendung der bekannten Vektorquantisierung wird dabei für jedes spezifische Geräusch in einer Trainingsphase ein Satz von "Verborgenen Markov-Modellen" erzeugt und abgespeichert. Beim Erkennen einer mit Geräusch behafteten Sprachäußerung muß dann eine Vielzahl von Vektorquantisierern vorgesehen werden, um die verschiedenen Markov-Modelle ansteuern zu können, damit dasjenige herausgefunden werden kann, welches die besten Erkennungsergebnisse liefert. Wie auf Seite 216 beschrieben wird, ist dazu eine sehr hohe Rechnerleistung erforderlich und es können nur Sprachäußerungen erkannt werden, die mit Geräuschen behaftet sind, für welche bereits Markov-Modelle vorhanden sind. Für den Fall, daß in dem zu erkennenden Eingangssignal ein Geräusch vorhanden ist, welches nicht katalogisiert ist, müssen zusätzliche Maßnahmen durchgeführt werden.

Zur Lösung dieses Problems ist eine zusätzliche sogenannte spektrale Subtraktion vorgesehen. Hierzu sind zusätzliche mit Vektorquantisierung trainierte Modelle von möglichen Geräushtypen erforderlich. Es wird dann die sukzessive Vektorquantisierung-Rahmenstörung über einen Tiefpaß gefiltert und ein Vergleich durchgeführt. Damit wird erreicht, daß ein möglichst ähnliches Geräusch erkannt wird, welches vom Eingabesignal nach dessen Vektorquantisierung subtrahiert wird. Das daraus resultierende Signal wird dann einem weiteren speziellen Vektorquantisierer zugeführt, womit eine reine Sprache repräsentiert werden soll, welche dann mit Markov-Modellen der reinen Sprache verglichen wird. Bei einem derartigen Verfahren ist eine große Speicherkapazität erforderlich, und die Rechnerleistung muß relativ hoch sein.

Ausgehend vom vorgenannten Stand der Technik besteht die Aufgabe der Erfindung darin, ein Verfahren anzugeben, mit dem bei relativ geringem Speicher und Rechneraufwand zunächst ein Geräusch eindeutig oder annähernd bestimmt werden kann. Mit dem dabei erzielten Ergebnis soll die Spracherkennung optimiert werden können.

Zur Lösung dieser Aufgabe ist eine Merkmalskombination vorgesehen, wie sie im Patentanspruch 1 angegeben ist.

Damit wird in vorteilhafter Weise erreicht, daß von jedem Störgeräushtyp lediglich eine Tabelle über die Häufigkeitsverteilung von Merkmalsvektoren gespeichert sein muß. Das Vergleichen einer vom aktuell eingegebenen Signal erstellten Häufigkeitsverteilungs-Tabelle der dabei ermittelten Merkmalsvektoren mit gespeicherten Vergleichstabellen nimmt wenig Rechnerleistung in Anspruch. Das beim Erkennen eines Störgeräushtyps auftretende Signal kann außer zur Spracherkennung auch für andere Schaltbefehle benutzt werden.

Ein Ausführungsbeispiel der Erfindung wird nachfolgend anhand von Zeichnungen näher erläutert.

Es zeigt Fig. 1 ein Blockschaltbild für das Trainieren und Erkennen von Störgeräushtypen

Fig. 2 ein Blockschaltbild einer Anordnung zur Erkennung von gestörter Sprache unter Verwendung von mehreren Codebüchern

In der Fig. 1 ist eine Anordnung dargestellt, die sowohl zum Trainieren von Sprach- und Geräuschmodellen dient und auch zum Erkennen von Sprache und Geräuschen geeignet ist. Um dies zu ermöglichen, wird zunächst ungestörte Sprache Spr von einem Mikrofon M aufgenommen und einer Sprachanalyseeinheit SA zugeführt. Dort werden einzelne Zeitabschnitte gleicher Länge in sogenannte Merkmalsvektoren MV überführt. Über einen Vektorquantisierer VQ werden aus diesen Merkmalsvektoren MV die Codebuchvektoren CV1 bis CVm für das Standardcodebuch SCB trainiert und dort eingetragen sowie jeweils mit einem Index I1 bis Im versehen.

Wenn auf diese Weise ein Standardcodebuch SCB mit ungestörter Sprache trainiert worden ist, so sind darin Codebuchvektoren CV1 bis CVm enthalten, die jeweils ihren zugeordneten Index I1 bis Im haben. Danach kann damit begonnen werden, Sprachäußerungen Spr zusammen mit Störgeräushtypen SG1 bis SGn über das Mikrofon M einzugeben. Die vom Mikrofon M aufgenommenen Signale werden von der Sprachanalyseeinheit SA, wie zuvor bereits beschrieben worden ist, in Merkmalsvektoren MV überführt und dem Vektorquantisierer VQ zugeführt. Der Vektorquantisierer VQ vergleicht nun alle Codebuchvektoren CV1 bis CVm mit dem gerade anstehenden Merkmalsvektor MV. Dabei wird herausgefunden, welcher Codebuchvektor CV1 bis CVm dem angebotenen Merkmalsvektor MV am ähnlichsten ist. Der dazugehörige Index, z. B. I2 gelangt dann in eine Erfassungseinrichtung EE und wird dort zwischengespeichert. Dies geschieht nacheinander mit allen Merkmalsvektoren MV, wobei in der Erfassungseinrichtung EE jedem Index zugeordnet die Häufigkeit seines Auftretens gespeichert wird. Auf diese Weise entsteht eine Zahlenreihe, die für jeden einzelnen Index I1 bis Im angibt, wie oft er während der Eingabe der mit einem Störgeräushtyp, z. B. SG1, behafteten Sprache Spr aufgetreten ist. Diese Zahlenreihe stellt eine Häufigkeitsverteilung der einzelnen Merkmalsvektoren MV dar, welche für den betreffenden Störgeräushtyp SG1 als charakteristische Häufigkeitsverteilung PR1 in einem Indexspeicher ISP eingetragen wird. Auch für die ungestörte Sprache Spr kann auf diese Weise eine Häufigkeitsverteilung PR0 im Indexspeicher ISP vorhanden sein. Der Indexspeicher ISP weist Bereiche, beispielsweise Zeilen auf, die jeweils einem Störgeräushtyp SG1 bis SGn und auch der ungestörten Sprache Spr zugeordnet sind, in denen jeweils die Häufigkeitsverteilungen PR0 bis PRn abgelegt sind.

Mit dem Inhalt dieses Indexspeichers ISP ist es nun möglich, einen Störgeräushtyp SG1 bis SGn zu erkennen. Zu diesem Zweck wird eine Sprachäußerung Spr zusammen mit einem unbekannten Störgeräushtyp SGu vom Mikrofon aufgenommen und der Sprachanalyseeinheit SA zugeführt. Die dabei entstehenden Merkmalsvektoren MV werden nun durch den Vektorquantisierer VQ mit allen im Standardcodebuch SCB vorhandenen Codebuchvektoren verglichen. Die dabei sich ergebenden Indizes I1 bis Im von dem jeweils ähnlichsten Code-

buchvektor CV werden in der Erfassungseinrichtung EE zwischengespeichert, wobei ermittelt wird, wie häufig jeweils ein Index I aufgetreten ist. Die dabei sich ergebende Häufigkeitsverteilung PT wird einem Vergleichsrechner VGR angeboten, welcher die Häufigkeitsverteilung PT mit allen im Indexspeicher ISP befindlichen Häufigkeitsverteilungen PR0 bis PRn vergleicht. Bei diesem Vorgang wird vom Vergleichsrechner ein Abweichungswert d ermittelt, der sich wie folgt berechnet:

$$d(P_R(\underline{x}), P_T(\underline{x})) = \sum_i P_T(x_i) \ln \frac{P_T(x_i)}{P_R(x_i)}$$

Wenn sich bei dieser Gegenüberstellung der Häufigkeitsverteilung PT von Indizes aus den aktuellen Merkmalsvektoren MV und den im Indexspeicher ISP befindlichen Häufigkeitsverteilungen PR0 bis PRn ein Abweichungswert d ergibt, der ein Minimum aufweist, so hat eine Störgeräuscherkennung SGE stattgefunden. Der Vergleichsrechner VGR stellt dabei fest, bei welcher im Indexspeicher ISP befindlichen Häufigkeitsverteilung PR0 bis PRn sich dieser minimale Abweichungswert d ergeben hat. Daraus resultiert die Beschaffenheit des Signals für die Störgeräuscherkennung SGE, welches angibt, um welchen Störgeräuschartyp SG1 bis SGn es sich handelt. Die Störgeräuscherkennung SGE kann in Form eines binär kodierten Wortes ausgegeben werden, welches direkt als Schalt- oder Ansteuerbefehl verwendet werden kann.

In der Fig. 2 ist eine Anordnung dargestellt, bei der mehrere Codebücher SCB, CB1 bis CBn verwendet werden. Diese Anordnung dient dazu, eine Spracherkennung zu optimieren, wenn bekannte oder unbekannte Störgeräusche vorhanden sind. Das Mikrofon M nimmt die Sprachäußerung Spr zusammen mit bekannten Störgeräuschen SG1 bis SGn oder unbekannten Störgeräuschen SGu auf und gibt sie an die Sprachanalyseeinheit SA ab. In der Sprachanalyseeinheit SA entstehen Merkmalsvektoren MV, die einem Vektorquantisierer VQ zugeführt werden. Der Vektorquantisierer VQ ist mit einem Codebuchmultiplexer CBM verbunden, welcher durch das Störgeräuscherkennungssignal SGE eingestellt wird. Damit wird gezielt dasjenige Codebuch SCB, CB1 bis CBn angesteuert, welches für den erkannten Störgeräuschartyp zuständig ist. Selbstverständlich müssen die für verschiedene Störgeräuschartypen SG1 bis SGn vorhandenen Codebücher CB1 bis CBn vorher trainiert worden sein. Für die ungestörte Sprache Spr ist das bereits beschriebene Standardcodebuch SCB vorhanden. Der Störgeräuschartyp SG0 entspricht demnach einer ungestörten Spracheingabe Spr. In einer Auswerteeinrichtung AE werden dann diejenigen Codebuchvektoren CV aus dem betreffenden Codebuch CB1 bis CBn oder auch SCB herausgefunden, welche den vom Eingangssignal gebildeten Merkmalsvektoren MV am ähnlichsten sind, um ein Spracherkennungssignal SE zu erhalten. Bei der in Fig. 2 dargestellten Anordnung sind selbstverständlich die funktionsgleichen Komponenten Sprachanalyseeinheit SA, Vektorquantisierer VQ sowie das Standardcodebuch SCB mit den in der Fig. 1 dargestellten gleichartigen Einrichtungen identisch.

Mit der in Fig. 2 dargestellten Anordnung ist es also möglich, für eine optimale Spracherkennung gezielt eines von mehreren vorhandenen Codebüchern SCB, SG1 bis SGn und den jeweils dazu gehörenden diskreten Markov-Modellen (HMMS, HMM1 bis HMMn) anzusteuern. Hierzu dient das Störgeräuscherkennungssignal SGE, welches den Codebuchmultiplexer CBM so einstellt, daß der Vektorquantisierer VQ mit einem der Codebücher SCB, CB1 bis CBn verbunden wird. Dabei können auch bei einem unbekannten Störgeräusch SGu, wofür kein Codebuch besteht, bessere Ergebnisse bei der Spracherkennung SE erzielt werden.

Das Störgeräuscherkennungssignal SGE kann auch dazu verwendet werden, die betreffende Störgeräuschquelle zumindest vorübergehend abzuschalten. So ist es beispielsweise möglich, auf Grund des erkannten Störgeräuschartyps SG1 bis SGn mit entsprechenden elektrischen oder elektronischen Schalteinrichtungen einen Lüfter auszuschalten oder ein Fenster bzw. eine Tür zu schließen. Das Störgeräuscherkennungssignal SGE kann auch dazu dienen, irgendwelche Einrichtungen zu überwachen, welche Geräusche von sich geben und dabei ihren Betriebszustand erkennen lassen. Es kann also erkannt werden, ob ein wichtiges elektrisches oder elektronisches Gerät in Betrieb ist oder nicht. Hieraus könnten Alarmmeldungen abgeleitet werden, wenn z. B. ein bestimmter Geräuschartyp von seinem Sollwert abweicht. Außerdem kann eine Spracherkennung SE noch weiter verbessert werden, wenn Störgeräuschartypen SG1 bis SGn eliminiert worden sind und eine nahezu ungestörte Sprache Spr ausgewertet werden kann.

#### Patentansprüche

1. Verfahren zum Ermitteln und Klassifizieren von Störgeräuschartypen unter Anwendung von für die Sprachanalyse bekannten Verfahren, wobei in einer Trainings- oder Lernphase Referenzmuster gebildet werden, die akustisch eingegebene Signale beschreiben und in einem Codebuch abgespeichert werden, um als Grundlage zur späteren Erkennung dieser Signale zu dienen, dadurch gekennzeichnet, daß zunächst ein Standardcodebuch (SCB) erstellt wird, welches die Referenzmuster von Sprachäußerungen (Spr), die ohne Störgeräusch (SG0) eingegeben und trainiert wurden, in Form von mit Indizes (I1 bis Im) versehenen Codebuchvektoren (CBV1 bis CBVm) enthält, daß danach die Sprachäußerungen (Spr), ohne und mit verschiedenen Störgeräuschartypen (SG0 bis SGn) überlagert, einzeln verarbeitet werden, wobei von den jeweiligen Merkmalsvektoren (MV), die durch Vektorquantisierung (VQ) mit dem Standardcodebuch (SCB) resultierenden Häufigkeitsverteilungen (PR0 bis PRn) der Indizes (I) durch eine Erfassungseinrichtung (EE) ermittelt und in einem Indexspeicher (ISP) abgespeichert werden, daß nach Abschluß der Trainingsphase von aktuell eingegebenen Sprachäußerungen (Spr), die mit einem

zunächst unbekannten Störgeräushtyp (SGu) behaftet sind, Merkmalsvektoren (MV) gebildet werden und die Häufigkeitsverteilungen (PT) der Indizes (I), die aus der Vektorquantisierung (VQ) mit dem Standardcodebuch (SCB) resultieren, ebenfalls ermittelt werden,

daß dann diese Häufigkeitsverteilung (PT) mit allen im Indexspeicher (ISP) abgelegten Werten der jeweils einem Störgeräushtyp (SG0 bis SGn) zugeordneten Häufigkeitsverteilungswerten (PR0 bis PRx) verglichen wird, indem in einem Vergleichsrechner (VGR) ein Abweichungswert (d) durch Summenbildung errechnet wird,

und daß der gesuchte Störgeräushtyp (z. B. SG1) dann gefunden ist, wenn dieser Abweichungswert (d) einen Extremwert hat.

2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß nach dem Ermitteln eines Störgeräushtyps (z. B. SG1) über einen Codebuchmultiplexer (CBM) gezielt ein dafür vorher trainierter Hidden-Markov-Spracherkenner mit an den Störgeräushtyp (SG1) angepaßten Codebuch (CB1) und Modellen (HMM1) angesteuert wird, wodurch eine genauere Spracherkennung (SE) ermöglicht wird.

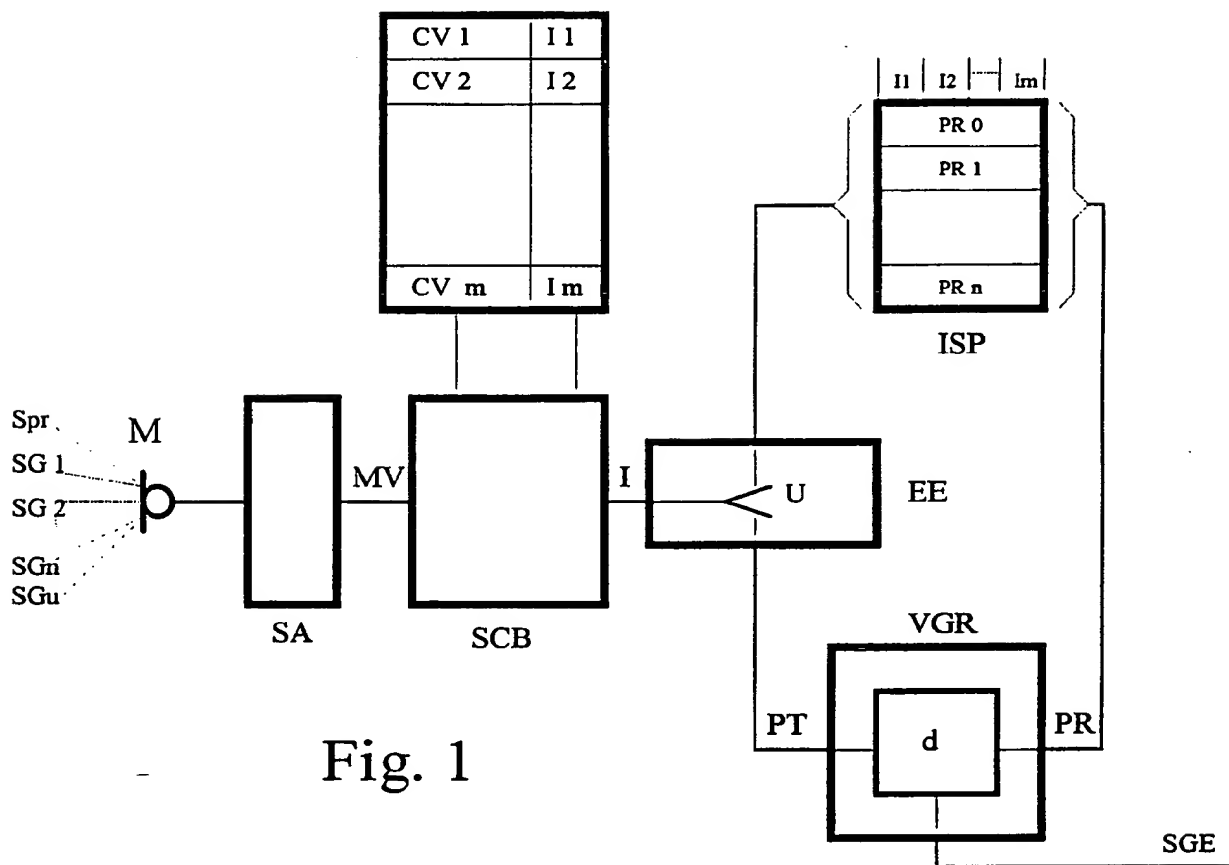
3. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß aufgrund des erkannten Störgeräushtyps (z. B. SG2) Schaltmaßnahmen durchführbar sind, die eine Beseitigung oder Reduzierung der Wirkung dieser Störgeräuschquelle bewirken können.

4. Verfahren nach Anspruch 3, dadurch gekennzeichnet, daß nach der Eliminierung eines Störgeräushtyps (z. B. SG2) erneute Spracheingaben und Berechnungen vorgenommen werden, wenn dies zur Ermittlung von weiteren Störgeräushtypen (SG) oder zur besseren Spracherkennung (SE) erforderlich ist.

---

Hierzu 2 Seite(n) Zeichnungen

---



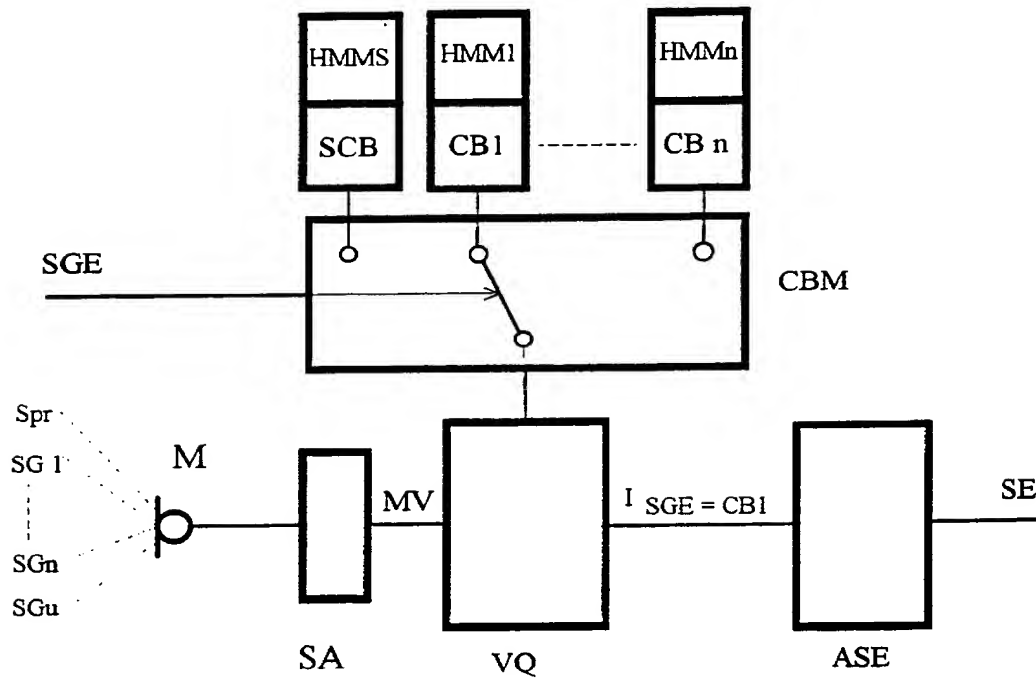


Fig. 2

TN P 4382